12

US ARMY
MATERIEL
COMMAND

TECHNICAL REPORT BRL-TR-2696

A NONPARAMETRIC STATISTICAL APPROACH TO THE VALIDATION OF COMPUTER SIMULATION MODELS

William E. Baker
Malcolm S. Taylor

November 1985

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

US ARMY BALLISTIC RESEARCH LABORATORY
ABERDEEN PROVING GROUND, MARYLAND

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (*When Data Entered*)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER
TECHNICAL REPORT BRL-TR- 2696

2. GOVT ACCESSION NO.
AD-A163376

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (*and Subtitle*)
A NONPARAMETRIC STATISTICAL APPROACH TO THE
VALIDATION OF COMPUTER SIMULATION MODELS

5. TYPE OF REPORT & PERIOD COVERED

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(*s*)
WILLIAM E. BAKER
MALCOLM S. TAYLOR

8. CONTRACT OR GRANT NUMBER(*s*)

9. PERFORMING ORGANIZATION NAME AND ADDRESS
US Army Ballistic Research Laboratory
ATTN: SLCBR-SE
Aberdeen Proving Ground, MD 21005-5066

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS
US Army Ballistic Research Laboratory
ATTN: SLCBR-DD-T
Aberdeen Proving Ground, MD 21005-5066

12. REPORT DATE
November 1985

13. NUMBER OF PAGES
40

14. MONITORING AGENCY NAME & ADDRESS(*if different from Controlling Office*)

15. SECURITY CLASS. (*of this report*)
UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (*of this Report*)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (*of the abstract entered in Block 20, if different from Report*)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (*Continue on reverse side if necessary and identify by block number*)

Computer Simulation Model	Wilcoxon Signed - Ranks Test
Validation	Mann - Whitney Test
Statistics	VAST Computer Simulation Model
Nonparametric	

the authors

20. ABSTRACT (*Continue on reverse side if necessary and identify by block number*)

Initially we completed a literature search in order to identify existing methods of computer simulation validation. Nonparametric statistical techniques were subsequently adapted to both deterministic and stochastic simulations, and these procedures were applied to a computer model currently in use at the Ballistic Research Laboratory. Monte-Carlo methods provided an indication of the power of these tests, and a mention of future work concerning attempts to increase this power has been included in this report.
Keywords: VAST (Vulnerability Analysis for Surface Targets).

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (*When Data Entered*)

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

I. INTRODUCTION

For three decades interest in simulation modeling and simulation languages has been expanding, almost keeping pace with the phenomenal rate of growth of computer technology. Lagging somewhat behind has been the concern for the validation of the resulting simulation models; that is, the establishment of some level of confidence that the model does, in fact, accurately mimic some real-world process. In the last fifteen years, research in validation techniques has been substantially increased; and a consensus of general conclusions has formed:

1. validation is problem dependent - there is no one general validation technique, mainly because the output from a model may be independent or correlated, univariate or multivariate, stationary or dynamic, and so forth; in fact, the model itself may be deterministic or stochastic,

2. in general, absolute validity is nonexistent - once a particular technique has been established, the model is usually validated only for a specific purpose and over a specific range of values,

3. empirical data are necessary - in order to validate a model, some comparison of output data with real-world data must be made; furthermore, these empirical data must be independent of those used in construction of the model, and

4. statistical tests are desirable - of the many methods proposed for validating simulation models, the use of statistical tests seems to be preferred, possibly because of the ability to establish some level of confidence.

Because computer simulation models are prevalent at the Ballistic Research Laboratory, the Experimental Design and Analysis Branch of the Systems Engineering and Concepts Analysis Division was funded to perform research in the area of the validation of such models. Results from the research are summarized in this report. They include a thorough literature review in which we examined existing validation techniques along with additional related information. Eventually we developed two nonparametric procedures, demonstrating them on a simulation model currently used by the Vulnerability/Lethality Division.

Nonparametric validation methods generally involve a procedure known as hypothesis testing. The initial step is to state a null hypothesis, usually "the simulation model is valid." Then a level of confidence is established, often 95%; and a particular test statistic is chosen. Two different errors are present in hypothesis testing. The first is called a Type I error and occurs when a true null hypothesis is rejected. If the level of confidence has been set at 95%, then it follows that the probability of a Type I error is 5%. However, in simulation model validation a Type II error is the more important to control; this occurs when a false null hypothesis is accepted. No level of confidence is pre-established to guard against accepting an invalid model; but, for any particular statistical test, a measure of the protection against this error is given by the power of the test, equal to the probability of rejecting the null hypothesis when it is false.

Unfortunately, there is a tradeoff between the two error types; as the level of confidence is increased (lower probability of a Type I error), the power of the test is decreased (higher probability of a Type II error). This implies that one way to increase the power of a test is to decrease the level of confidence in it. There are, however, more satisfactory ways; and they will be mentioned in the summary of this report. The important point to remember is that when attempting to validate a simulation model using hypothesis testing, it is imperative that the statistical test be a powerful one.

II. LITERATURE REVIEW

As the electronic computer became a more powerful tool, computer simulation became a more viable method by which the behavior of a given process could be characterized. As early as the 1950's, articles were being published about computer modeling of entire systems; and soon after, specialized simulation languages were developed. The pioneers in this field realized the need for some assurance that the simulation output would be consistent with the empirical data that were available. However, prior to 1967 there was very little written that provided any explicit procedures which might be applied to determine the soundness of a computer model. In that year several papers concerning this problem were published, and two of them became a foundation upon which most subsequent efforts have been constructed.

In 1967, Fishman and Kiviat[1] provided definitions which differentiated the notions of verification and validation, terms which had previously been used interchangeably. "Verification determines whether a model with a particular mathematical structure and data base actually behaves as an experimenter assumes it does. Validation tests whether a simulation model reasonably approximates a real system." Most individuals working in this area today have subscribed to these definitions, although papers continue to be published which do not discriminate between the two ideas. Figure 1, taken from a paper by Winter, et. al.[2], is a Venn diagram illustrating the relationship between verification, validation, and other concepts within the field of computer simulation. Stone[3] believed the word assessment "... is preferable to validation which has a ring of excessive confidence about it." However, in this paper we will continue to consider validation as defined by Van Horn,[4] who expanded on the previous definition by giving it a somewhat statistical flavor. "Validation ... is the process of building an acceptable level of confidence that an inference about a simulated process is a correct or valid inference for the actual process."

[1] Fishman, G.S. and Kiviat, P.J., "Digital Computer Simulation. Statistical Considerations," Memorandum RM-5387-PR, The Rand Corporation, 1967.

[2] Winter, E.M., Wisemiller, D.P., and Ujihara, J.K., "Verification and Validation of Engineering Simulations with Minimal Data," Proceedings of the 1976 Summer Computer Simulation Conference, 1976.

[3] Stone, M., "Cross-Validating Choice and Assessment of Statistical Prediction," Journal of the Royal Statistical Society, series B-36, 1974.

[4] Van Horn, R., "Validation," The Design of Computer Simulation Experiments, Duke University Press, 1969

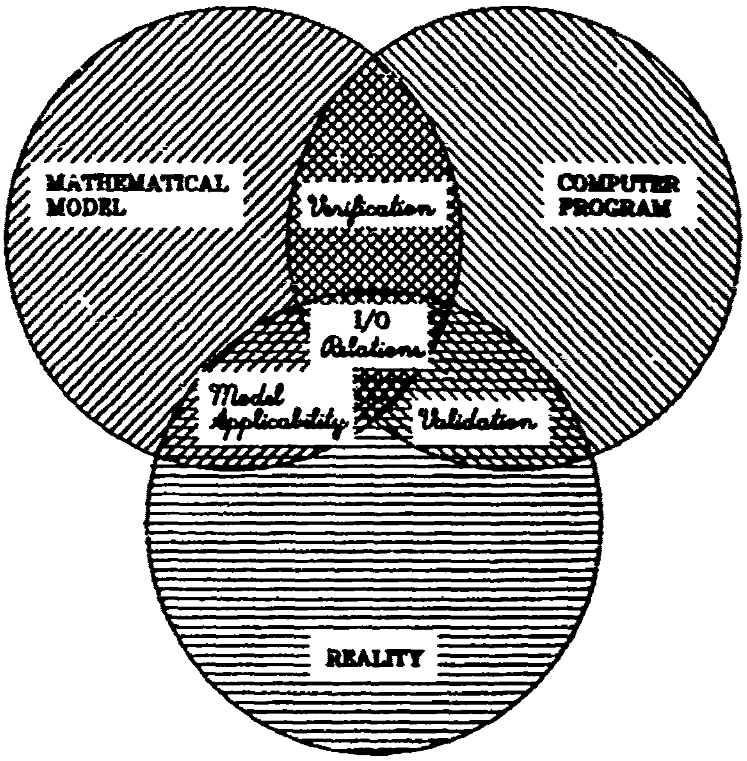

FIGURE 1: RELATIONSHIPS BETWEEN THE VARIOUS CONCEPTS OF A COMPUTER SIMULATION

The second influential paper to appear in 1967 was by Naylor and Finger.[5] In it they proposed a three-stage approach to validation of a computer simulation. This technique, or a modified version of it, has been used by numerous authors. Law[6] has augmented their approach with specific suggestions for each of the three stages:

1. develop high face-validity - insure that the simulation seems reasonable to those people who are knowledgeable in the area,

2. test the simulation assumptions - examine the data used in building the simulation and empirically test the assumptions drawn from those data, and

3. compare simulation output data with empirical data - use tests, statistical if possible, to determine a level of confidence in the simulation.

When attempting to validate existing models, the first two stages will often have already been completed by the developer of the simulation leaving only the third stage, potentially the most difficult.

[5] Naylor, T H and Finger, J M , "Verification of Computer Simulation Models," Management Science, Vol 14 No 2, 1967

[6] Law, A M , Simulation Modeling and Analysis, University of Wisconsin, 1979

Not everyone subscribes to the three-stage approach to validation. However, there does seem to be a general agreement that the third stage, comparing simulation output data with empirical data, is crucial. Sometimes obtaining empirical data in the region of applicability is very difficult, especially in engineering simulations. Winter, et. al.[2] mention in that case, "The quality of the component models and the excellent knowledge of the random process along with a systematic verification must be a substitute for validation." However, Fishman and Kiviat[1] are firm in their statement that " ... if no numerical data exist for an actual system, it is not possible to establish the quantitative congruence of a model with reality." In attempting to perform this third stage, Wright[7] suggests that three questions be considered:

1. how do we intelligently compare simulation output data with empirical data,

2. how do we collect and exploit the empirical data used in our tests, and

3. how do we transform the results of these tests into a confidence in the computer simulation?

Finally, Baird, et. al.[8] warn that the empirical data used for comparison with the simulation output data must be independent of those used in building the computer model; otherwise, we have only verification of the simulation.

Tytula[9] has divided the many methods used for the data comparison into five general categories:

1. judgemental comparison - this method seems to be the most widely used and includes graphical analysis and the comparison of common properties such as the mean and variance; it is easy to use and quite practical, but the impact of errors in judgement is difficult to assess,

2. hypothesis testing - this method includes goodness-of-fit tests, analysis-of-variance techniques, and nonparametric ranking methods; since this will be the category of interest in our report, the advantages and disadvantages will be discussed in the succeeding section,

3. spectral analysis - since the output of many simulation models is in the form of a time series, this method is particularly useful; however, it is difficult to relate the invalidity at a particular frequency to the overall simulation validity,

[7] Wright, R.D., "Validating Dynamic Models: An Evaluation of Tests of Predictive Power," Proceedings of the 1972 Summer Computer Simulation Conference, 1972

[8] Baird, A.M., Goldman, R.B., Bryan, W.C., Holt, W.C., and Belrose, F.M., "Verification and Validation of RF-Environmental Models - Methodology Overview," Boeing Aerospace Company, 1980

[9] Tytula, T.P., "A Method for Validating Missile System Simulation Models," Technical Report E-78-11, U.S. Army Missile Research and Development Command, 1978

4. sensitivity analysis - this method can determine a range of parameter values and assumptions over which the simulation is valid, but it is usually difficult to analyze the effects of the characteristics drifting outside this range, and

5. indices of performance - this method is useful in ranking models; however, it is impossible to pick a value for a given index which will always imply a valid simulation.

Validation is a difficult process because, as Tytula[9] points out, no single satisfactory method exists. Most techniques are problem dependent; and, indeed, the output data of a simulation may be independent or correlated, univariate or multivariate, stationary or dynamic. In fact, Garrett[10] states that, "The critical dimension affecting the applicability of various techniques is that of the deterministic or stochastic nature of the output." Only a few authors have attempted to provide a general validation technique - see Gilmour[11] for an example. Most have developed methods which apply to a select subset of simulation models; and, even then, the simulation is often validated only for a particular purpose or over a particular range of values. In tha' case, care must be taken not to apply the simulation model outside the validated regio.

III. VALIDATION PROCEDURES

In this report we will be examining hypothesis testing as a method for validating both deterministic and stochastic computer simulation models. This type of procedure allows some level of confidence to be attached to the results. When employing hypothesis testing, several assumptions must usually be stated; but by using nonparametric ranking techniques we will eliminate one major (and often unjustifiable) assumption - that the data arise from a normal distribution.

Sargent[12] notes that for hypothesis testing we generally assume a null hypothesis that the simulation model is valid. Then by establishing a level of confidence for a particular statistical test, we fix the probability of a Type I error in which we reject a valid model. However, for simulation validation it is more important to minimize the probability of a Type II error, that is, accepting an invalid model. The magnitude of the Type II error can be determined by the power function of the statistical test where the power is the probability of rejecting a false null hypothesis. For a fixed sample size there is a tradeoff between the two error types, so that we can increase the power at the expense of the confidence level. Unfortunately, the power can not be computed against

[10] Garrett, M., "Statistical Validation of Simulation Models," Proceedings of the 1974 Summer Computer Simulation Conference, 1974

[11] Gilmour, P., "A General Validation Procedure for Computer Simulation Models," The Australian Computer Journal, Vol 5 No 3, 1973

[12] Sargent, R G., "Developing Statistical and Cost-Risk Procedures for Validation of Simulation Models," U S Army Research Office Proposal Number 18201-M, 1980

an alternative hypothesis as general as, "The simulation model is invalid"; and therefore, it must be examined against an array of different specific alternative hypotheses. Nevertheless, we continue to search for powerful statistical tests with justifiable assumptions which will still provide acceptable levels of confidence.

Let $X = (x_1, x_2, ..., x_k)$ be a vector of inputs to a simulation model, and let y be an output resulting from X. Then y may take on a single value, as in a deterministic model, or many values, as is the case with a stochastic model. Let z be the corresponding value from the real-world process given the same input vector. In general, y will not be equal to z since X contains only a finite number of input variables; ostensively, the most relevant ones. The purpose of the simulation model is to mimic the real-world process. Thus, in attempting to validate it, we compare each empirical value with the corresponding model output generated under the same conditions; that is, the same values for the vector X.

Suppose there exist N pairs of data $(y_1, z_1), (y_2, z_2), \ldots, (y_N, z_N)$ available for comparison, where each pair corresponds to a different input vector and where each y_i may itself be a vector of values in the case of a stochastic model. Reynolds and Deaton[13] note that because each of the pairs was generated under different conditions, it would be incorrect to pool the data and proceed with the testing of our hypothesis. Rather, we must find a statistical procedure which examines each pair individually and then allows for the combination of these results into one overall test that provides reasonable power. With this as our goal, we propose to use two nonparametric statistical procedures - the Wilcoxon signed-ranks test in the case of a deterministic model and, for a stochastic model, a process which combines independent cases of the Mann-Whitney test.

Deterministic Model

A deterministic model provides one and only one set of output values for each set of input values. Such a model is frequently used as a first attempt at representing a stochastic system, and quite often it will adequately simulate at least the coarse behavior of such a system. The deterministic model generally has the advantages of being both simple and inexpensive. Any individual output value y from the model can be compared with an empirical value z obtained from the actual system under the same set of input values. Considering N different input sets, the available data consist of N observations $(y_1, z_1), (y_2, z_2), \ldots, (y_N, z_N)$ of bivariate random variables. The Wilcoxon signed-ranks test is applicable. The null hypothesis of this test can be loosely stated as, "The values of the y_i's tend to be the same as the values of the z_i's," which we can interpret as, "The simulation model is valid."

[13] Reynolds, M.R., and Deaton, M.L., "Comparisons of Some Tests for Validation of Stochastic Simulation Models," Commun. Statist. - Simula. Computa., Vol 11 No 6, 1982

The Wilcoxon signed-ranks tests is a hypothesis test for identical medians that uses paired observations. To use it, we first compute $D_i = y_i - z_i$ for $i = 1, 2, ..., N$, recalling that each of these random variables may be from a different distribution. The following four assumptions are made concerning these D_i's:

1. the distribution of each D_i is symmetric,

2. the D_i's are mutually independent,

3. the D_i's all have the same median, call it $m_{.50}$, and

4. the measurement scale of the D_i's is at least interval.

The fourth assumption means that for any two observations on the random variable we can distinguish not only which is larger and which is smaller, but also which is farther from the common median.

The null hypothesis is that $m_{.50} = 0$; in other words, that all the D_i's have medians equal to zero. This would indicate that the [illegible] of the y_i's and the z_i's tend to be the same. A rank R_i, based on the absolute value of each D_i, is assigned; thus, the R_i's consist of the integers 1 to N. R_i is then adjusted to zero for each $D_i < 0$. The non-zero integers that remain are the ranks of the positive D_i's; and a test statistic T is defined to be their sum; that is, $T = \sum_i R_i$. Very high and very low values of T cause rejection of the null hypothesis. The theory behind the test is explained very clearly by Conover[14], where tables containing various quantiles of the Wilcoxon signed-ranks test statistic are available.

One further assumption is sometimes made, that each D_i is a continuous random variable. Theoretically, this assures that there will be no $D_i = 0$ and no $D_i = D_j$ where $i \neq j$. However, in practice the available data may produce zeros and ties; and methods have been devised for handling these situations. Although it is often recommended that the zeros be dropped from the data immediately, they are sometimes very important, especially when attempting to show that there is no significant difference between the values of the y_i's and z_i's. Lehmann[15] proposes ranking the absolute values of all the D_i's including the zeros and, in the case of ties, assigning each of the tied values the average of the ranks normally due them. Then the R_i's are adjusted by multiplying them by -1 if $D_i < 0$, 0 if $D_i = 0$, or 1 if $D_i > 0$. The test statistic T_1 then becomes the sum of the positive R_i's, and a second test statistic T_2 is defined as the sum of the absolute values of the negative R_i's. Rejection of the null hypothesis is caused by very high values of either T_1 or T_2.

[14] Conover, W J., Practical Nonparametric Statistics, John Wiley & sons, Inc , 1971.

[15] Lehmann, E.L , Nonparametrics. Statistical Methods Based on Ranks, Holden-Day, Inc , 1975

As mentioned earlier, a misuse of hypothesis testing as a method of simulation validation occurs when too little concern is shown for the power of the test. The power is the probability of rejecting an invalid model, and we would like this probability to be as close to one as possible. Unfortunately, the power can be calculated only for specific alternative hypotheses. In order to generate power curves for the Wilcoxon signed-ranks test, it is convenient to make the additional assumption that all D_i's come from a common distribution. Although this may not always be valid, it does afford us an indication of the power of the test against an alternative consisting of a shift in the mean, which fo· a symmetric distribution is identical to the median. Figure 2 shows some power curves for this test against a shift in the mean when the underlying distribution of the D_i's is normal with a mean equal to μ and a variance equal to one. Recall that a true null hypothesis would indicate that the values of the y_i's and the z_i's tend to be equal. These curves were generated using a Monte-Carlo procedure which incorporated 10,000 replications. Note the increase in power as the number of observations increases. Figures 3-5 display some power curves for other alternative hypotheses, each figure assuming a different common distribution for the D_i's with a corresponding modification of one of the parameters of the distribution. Notice when the abscissa is equal to zero (when the null hypothesis is true), the probability of rejection is 0.05 - the value chosen for the probability of a Type I error. The faster the curve approaches one, the more powerful the test against that particular alternative hypothesis. Although very narrow in their scope, these results do provide us with an indication of the overall power of the test against a shift in location and allow us to determine the extent to which the probability of a Type II error might be reduced by an increase in sample size.

Stochastic Model

A stochastic model provides a set of output values that, for each given set of input values, occurs with a certain probability. Mihram[16] states that this "... probability ... serves as a measure of our human ignorance of the actual situation and its implications." Generally, the behavior of the system is too complicated to include all of the appropriate inputs in the computer model. Even if it were possible, the return in accuracy provided by such thoroughness may be small. Refinement of a computer model usually leads to stochastic modeling; and because of the abilities of today's computers, the use of such modeling has substantially increased.

Given M replications, output of the model becomes a set of values $y^1, y^2, ..., y^M$ for each set of input values which can be compared with (in our case) a single corresponding empirical value z. Recall that X is a vector of most, but not all, of the relevant input variables. Then z, given the value of X, is a random variable reflecting the random error due to the exclusion of certain factors from X. Also y, of course, is a random variable since the simulation model is stochastic. We would like to show that $F(y|X)$, the conditional distribution function of y, is equal to $G(z|X)$, the conditional distribution function of z for all $-\infty < y, z < \infty$ and for all X.

[16] Mihram, G A , Simulation. Statistical Foundations and Methodology, Academic Press, Inc , 1972

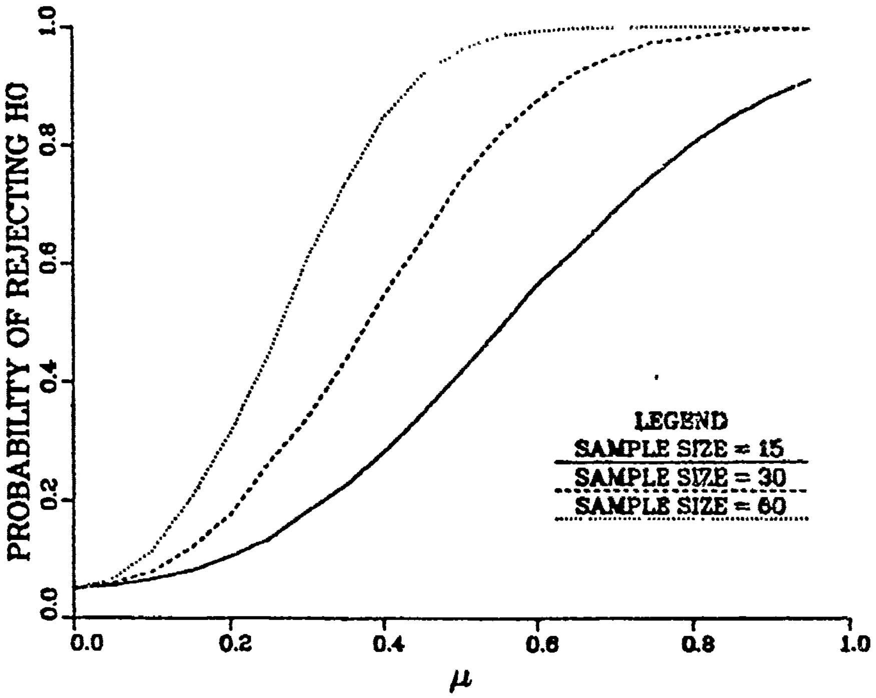

FIGURE 2: POWER OF 5%-LEVEL TEST
H0: F=NORMAL(0,1) VS. H1: F=NORMAL($\mu \neq 0$,1)

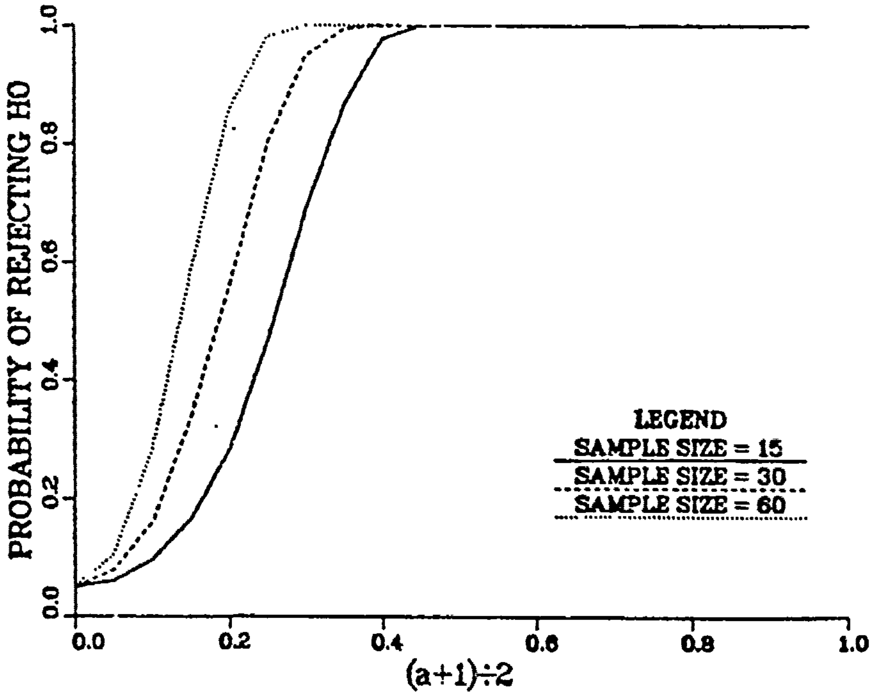

FIGURE 3: POWER OF 5%-LEVEL TEST
H0: F=UNIFORM(−1,1) VS. H1: F=UNIFORM(a≠−1,1)

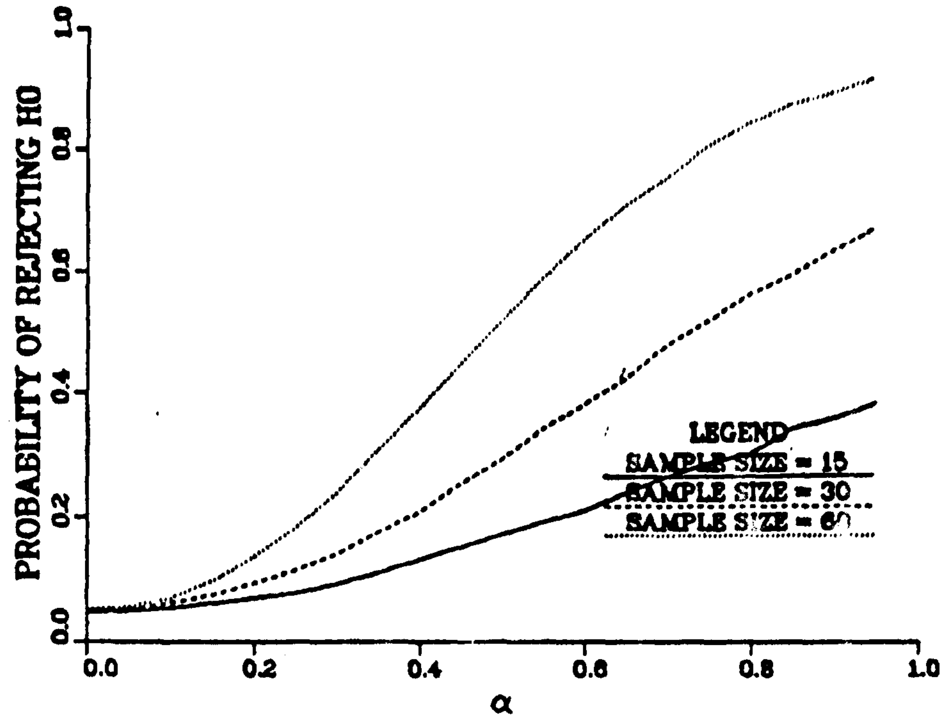

FIGURE 4: POWER OF 5%-LEVEL TEST
H0: F=CAUCHY(0,1) VS. H1: F=CAUCHY(α≠0,1)

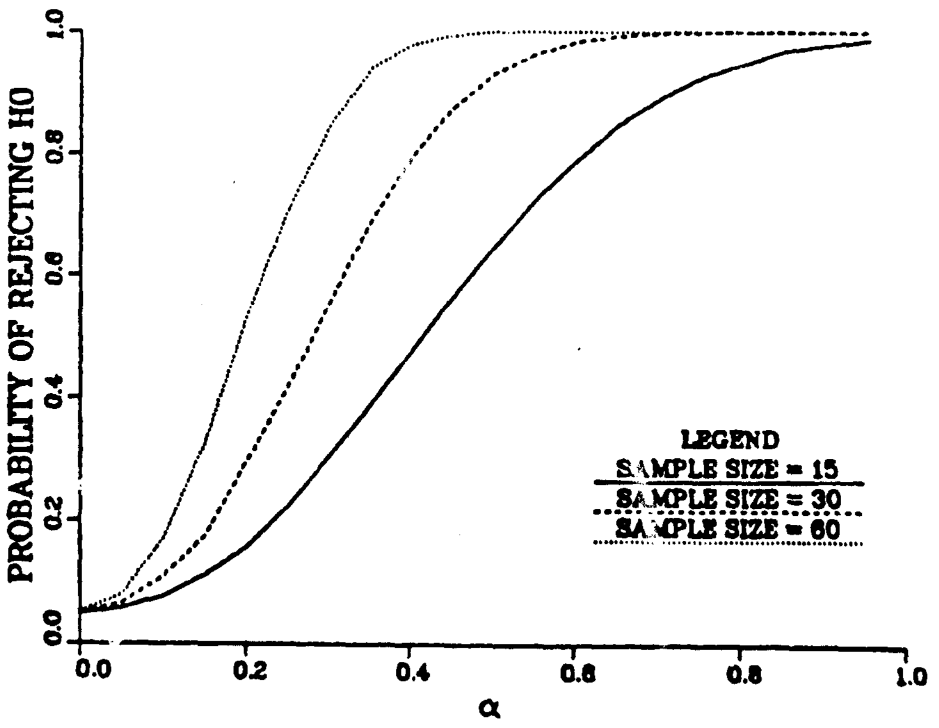

FIGURE 5: POWER OF 5%-LEVEL TEST
H0: F=LOGISTIC(0,1) VS. H1: F=LOGISTIC(α≠0,1)

Considering N different input sets, the available data consist of N observations $(y_1^1, y_1^2, ..., y_1^M, z_1)$, $(y_2^1, y_2^2, ..., y_2^M, z_2)$, . . ., $(y_N^1, y_N^2, ..., y_N^M, z_N)$ of multivariate random variables, where the y^k's for any given observation share a common distribution. Mihram[16] suggests ranking $y_i^1, y_i^2, ..., y_i^M, z_i$ for each i; if the model is valid, we would expect the z_i to fall somewhere in the middle of such a ranking. This is the initial step in a procedure known as the Mann-Whitney test, a particular case in which one of the random variables, namely z_i, has a sample size of one. Since we are dealing with N observations, we need a method by which we can combine independent cases of the Mann-Whitney test; such a method has been proposed by Van Elteren[17] and referenced in a very clear example by Reynolds, et.al.,[18].

The Mann-Whitney test is a hypothesis test involving samples from two distributions that tests for equality of the distributions. For each input set X a sample of M output sets $y^1, y^2, ..., y^M$ is obtained from the computer simulation, and the empirical observation z provides another sample of size one. The following three assumptions are made:

1) both samples are random samples from their respective populations,

2) in addition to independence within each sample, there is mutual independence between the two samples, and

3) the measurement scale is at least ordinal.

The third assumption means that for any two observations on the random variable we can distinguish which is larger and which is smaller.

The null hypothesis is that $F(y|X) = G(z|X)$ for a given input set X. When we combine N of these tests, in the manner suggested by Van Elteren, we have the null hypothesis of $F(y|X) = G(z|X)$ for all $-\infty < y, z < \infty$ and for all X, which we can interpret as, "The simulation model is valid." Let R_i be the rank of z_i in the i^{th} observation $(y_i^1, y_i^2, ..., y_i^M, z_i)$; thus, R_i is an integer between 1 and M + 1. Then a test statistic T is defined as the sum of the R_i's over all N observations; that is, $T = \sum_i R_i$. Very high or very low values of T will cause rejection of the null hypothesis. The theory behind the Mann-Whitney test is given in Conover[14], and the combination of such tests is explained by Van Elteren[17].

A fourth assumption is usually made, that both samples consist of random variables from continuous distributions. As in the case of the Wilcoxon test statistic, this is to assure that there will be no zeros and, more importantly, no ties. However, for

[17] Van Elteren,P., "On the Combination of Independent Two Sample Tests of Wilcoxon," Bulletin de l'Institute International de Statistique, 37, 1960

[18] Reynolds, M.R., Burkhart, H.E., and Daniels, R.F., "Procedures for Statistical Validation of Stochastic Simulation Models," Forest Science, Vol 27 No 2, 1981.

this test, a moderate number of ties is tolerable; and they are handled as previously by assigning each of the tied values the average of the ranks normally due them.

The power of this test against alternative hypotheses analogous to those shown for the Wilcoxon test is displayed in Figures 6-9 which were generated using a Monte-Carlo procedure which incorporated 2,000 replications. Once again, in generating these power curves, we have made one additional, albeit restrictive, assumption; namely, the distribution of the y_i's is the same for each vector of input values, and similarly for the distribution of the z_i's. Although it would be preferable to avoid this assumption, it is necessary in order to test against specific alternative hypotheses - in this case, a shift in the mean; and, as with the Wilcoxon test, these curves do provide an indication of the overall power of this combination of Mann-Whitney tests against the shift in location. This test appears slightly less powerful than the Wilcoxon signed-ranks test. This is a result of the assumption of the less stringent ordinal measurement scale. If $M = 1$, the combined Mann-Whitney test reduces to the sign test, a nonparametric procedure similar to the Wilcoxon test but making no assumption of symmetry of the distributions and consequently requiring only an ordinal measurement scale, resulting in a less powerful test. Reynolds and Deaton[13] look at some test statistics similar to T designed to be more powerful against other alternative hypotheses.

IV. EXAMPLE

The Vulnerability Analysis for Surface Targets (VAST) model is a computer simulation currently in use at the Ballistic Research Laboratory to evaluate the effect of kinetic energy projectiles or shaped-charge threats against a single surface target.[19] It incorporates damage from both the primary penetrator and any associated spall fragments; but currently it is unable to handle damage resulting from blast, heat, and certain synergistic effects such as ricochets. Furthermore, there is a variety of opinions, estimates, and decisions, all based on the experience of the vulnerability analysts but generally providing vague and imprecise data, which subsequently serve as input to the simulation. Nevertheless, results demonstrate reasonable face validity, so an attempt at statistical validation of the model seems feasible.

A target description is produced by a separate computer code using a combination of geometric figures and, once generated, can be viewed from any orientation. After a viewing angle has been established, a rectangular grid is superimposed over the target in the plane orthogonal to that angle. From a (uniform) randomly-selected point within each grid cell, a ray is traced through the target; and a list is constructed of all components encountered. If a spall-producing component is encountered, spall rays are traced from that point of impact to all critical components in the target. These rays represent spall fragments whose size, shape, and velocity are chosen at random from specified distributions.

[19] Hafer, T F and Hafer, A S , "Vulnerability Analysis for Surface Targets (VAST) An Internal Point-Burst Vulnerability Model," ARBRL-TR-02154, U S Army Ballistic Research Laboratory, 1979.

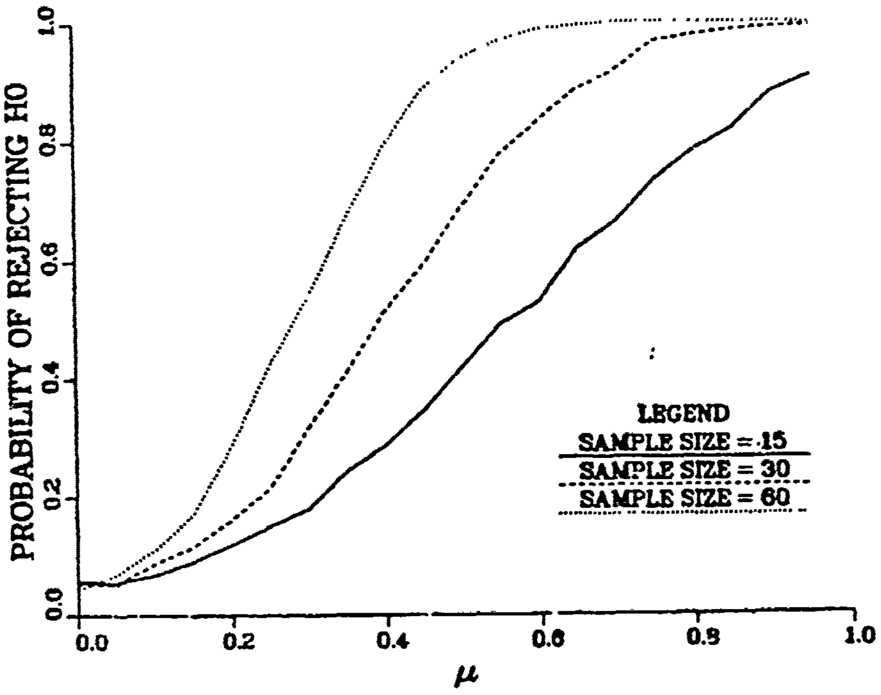

FIGURE 6: POWER OF 5%-LEVEL TEST
H0: F=G=NORMAL(0,1) VS. H1: F=NORMAL(0,1), G=NORMAL($\mu \neq 0$,1)

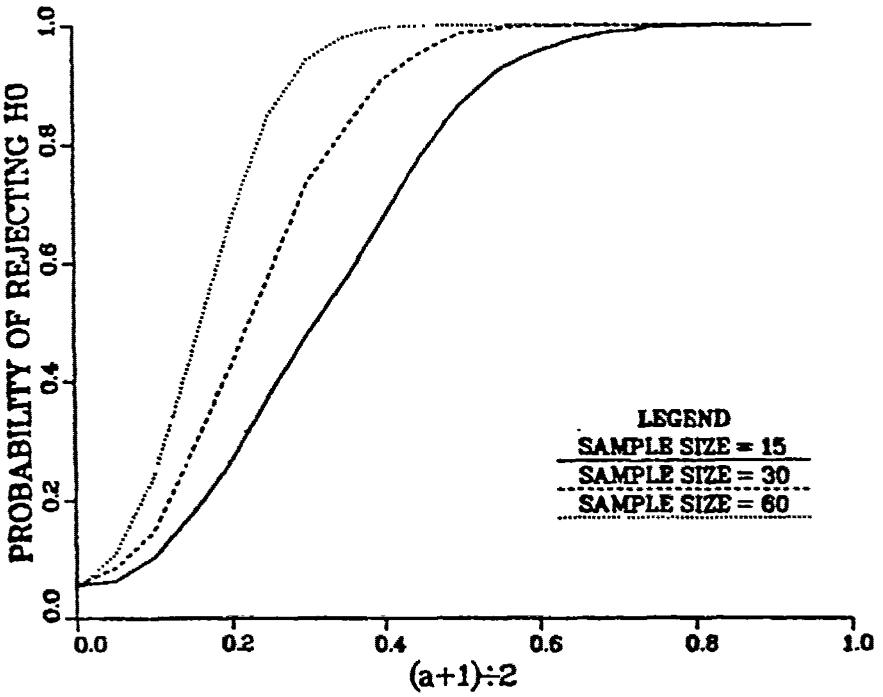

FIGURE 7: POWER OF 5%-LEVEL TEST
H0: F=G=UNIFORM(−1,1) VS. H1: F=UNIFORM(−1,1), G=UNIFORM($a \neq -1$,1)

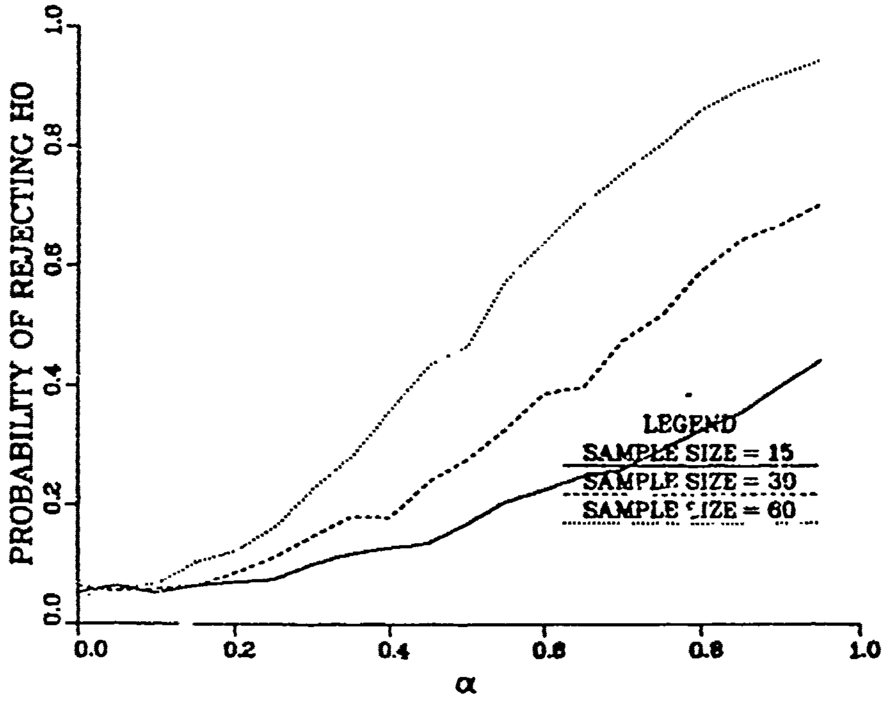

FIGURE 8: POWER OF 5%-LEVEL TEST
H0: F=G=CAUCHY(0,1) VS. H1: F=CAUCHY(0,1), G=CAUCHY(α≠0,1)

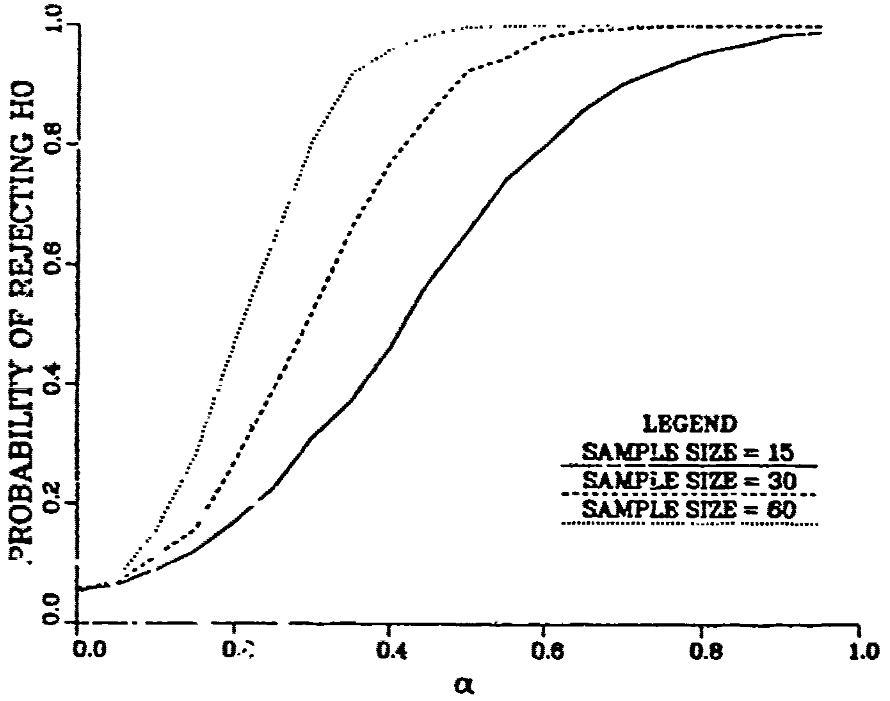

FIGURE 9: POWER OF 5%-LEVEL TEST
H0: F=G=LOGISTIC (0,1) VS. H1: F=LOGISTIC(0,1), G=LOGISTIC(α≠0,1)

Along each individual ray, residual masses and velocities of the primary penetrator and associated spall fragments are used to calculate the probability of incapacitation for each critical component. These are then combined over all critical components and provide a loss of function (LOF) for the particular cell, further combined over all cells to provide a LOF for the particular orientation, and finally combined over several orientations to provide an overall LOF for the target. Although its input is stochastic in nature, the VAST model is generally run with just one replication because the results are fairly consistent from replication to replication and because the model requires considerable time and, hence, expense to execute.

Data were provided by vulnerability assessors who had estimated loss of function for a particular surface target based on their inspection of actual damage from a particular round of ammunition - in this case, the function evaluated was the mobility function. When attempting to compare model output with this empirical data, it was first necessary to determine the exact point of impact on the surface target during the live-fire exercise. Then the VAST model assumed that point of impact to be the origin of the ray representing the primary penetrator. Damage due to that ray and its associated spall rays were then combined to provide a loss of function value which could be compared with the empirical datum point. Therefore, only one orientation was considered and, for that particular orientation, a ray originating at a specific point within only one cell was examined. Encountering a spall-producing component still required a random selection of spall characteristics; and because execution time was reduced, the model was run using thirty replications - the output data appear in Table 1. The averaged results were compared with the empirical data, in the manner proposed for deterministic simulations; individual outputs from the thirty two replications were also compared with the empirical data, this time using the method proposed for stochastic simulations. Thus, these data provided examples for both of our proposed validation procedures.

Results of the test for the deterministic form of the model appear in Table 2. Under the null hypothesis of a valid model, the sum of the positive ranks should equal the sum of the absolute values of the negative ranks; that is, $T_1 = T_2$. Lehmann[15] shows how to establish critical values against which the test statistic can be evaluated. He derives the expectation of the test statistic,

$$E[T] = \frac{1}{4}[N(N+1) - d_0(d_0+1)], \tag{1}$$

and the variance of the test statistic,

$$\mathrm{Var}[T] = \frac{1}{24}[N(N+1)(2N+1) - d_0(d_0+1)(2d_0+1)] - \frac{1}{48}[\sum_{i=1}^{n} d_i(d_i-1)(d_i+1)], \tag{2}$$

where T is either the sum of the positive ranks or the sum of the absolute value of the negative ranks, N is the number of observations, d_0 is the number of zero differences, and d_i represents the number of tied differences for the i^{th} tie with n different ties. Appealing to the central-limit theorem, $T^* = (T - E[T])/\sqrt{\mathrm{Var}[T]}$ tends to the

TABLE 1. LOSS OF FUNCTION VALUES - MOBILITY KILL

	Replications														
Shot Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
43	.7385	.7806	.6500	.8482	.6500	.7053	.7671	.6500	.6500	.6870	.8793	.6500	.8173	.7670	.7669
44	1.0000	1.0000	1.0000	.1000	1.0000	1.0000	.1000	1.0000	1.0000	.1000	1.0000	.1000	.1000	1.0000	1.0000
45	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
46	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.000	1.0000	1.0000	1.0000	1.0000
47	.1484	.1000	.1148	.1713	.1000	.1169	.1320	.1000	.1482	.1000	.1000	.1305	.1185	.1000	.1000
48	.8380	.6500	.7879	.7069	.8408	.7804	.7671	.6911	.8755	.6500	.8575	.7113	.7440	.7122	.7670
49	.6500	.6500	.6500	.6500	.6849	.6500	.6500	.6500	.6500	.6906	.6500	.6500	.6500	.6903	.6500
50	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
51	.8733	.6315	1.0000	.7989	.9205	1.0000	.9548	.8020	.8058	.9772	1.0000	.6316	.9995	1.0000	.9996
52	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
53	.4005	.9990	.3902	.4725	.4002	.9991	.6031	.4003	.3913	.4003	.3910	.6243	1.0000	.4597	.4003
54	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
55	.8912	.9424	.8465	.9046	.6870	.8679	.9065	.8789	.9166	.9648	.8884	.9184	.8988	.8156	.8839
56	.3000	.3937	.3000	.0134	.3049	0.0000	.3000	.3050	.3097	.3000	.3048	.3000	.3099	.3053	.3099
57	.6444	.2423	.1000	.6500	.1000	.6498	.6499	.7244	.7032	.6500	.6500	.7039	.6648	.6500	.6500
58	1.0000	.9989	1.0000	1.0000	.9998	.9866	1.0000	1.0000	1.0000	12.0000	.9999	.9998	.6580	1.0000	1.0000
59	.0059	.0512	.0500	.0500	.0487	.0060	.0489	.0690	1.0000	.0500	.0227	.0500	.0500	1.0000	.0521
60	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
62	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
64	.1000	.4015	.1000	0.0000	.0999	.1000	.1000	.1000	.1000	.1000	.3350	.0921	.1000	.5308	.1000
65	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
66	.8326	.8002	.7185	.7902	.8319	.7161	.7533	.6646	.8425	.6971	.6715	.7674	.7586	.8826	.7528
67	.8200	1.0000	1.0000	.9200	1.0000	1.0000	1.0000	.8960	.9299	1.0000	1.0000	1.0000	.9530	.7581	1.0000
68	.8945	.5497	.9978	.8978	.8020	.7220	.1000	.7197	.5500	.6486	.8942	.8995	.6498	.8982	.8937
69	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
70	.9997	1.0000	1.0000	1.0000	1.0000	1.0000	.9918	.9999	.6987	.6500	1.0000	.9765	.8996	1.0000	1.0000
71	.1000	.1337	.1628	1.0000	1.0000	1.0000	1.0000	1.0000	.2551	1.0000	.1000	.1000	.1000	.1000	1.0000
72	.9859	.9979	.8519	.9542	.9918	.9740	.9877	.9538	.9929	.8442	.9561	.9968	.8705	1.0000	1.0000
73	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
74	.6968	.4015	.8521	.4014	.4015	.6609	.6813	.8774	.4015	.6925	.8356	.6909	.6574	.6960	.4015
75	.6374	.6465	.6854	.6757	.6229	.5580	.6254	.6805	.8372	.2984	.6761	.6499	.7939	.6269	.7078
76	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

TABLE 1. LOSS OF FUNCTION VALUES - MOBILITY KILL (Cont'd)

Shot Number	Replications															
	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	Average
43	.6500	.7543	.6500	.6722	.7725	.6500	.6500	.7546	.6500	.6666	.7639	.7508	.6500	.7670	.7665	.719
44	.1000	1.0000	1.0000	1.0000	.1000	.1000	1.0000	.1000	1.0000	.1000	1.0000	1.0000	1.0000	1.0000	1.0000	.700
45	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.000
46	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.000
47	.1149	.1000	.1000	.1000	.1181	.1180	.1181	.1000	.1000	.1000	.1000	.1370	.1507	.1328	.1181	.116
48	.9885	.7197	.9895	.9903	.7127	.7258	.7673	.7672	.6716	.6500	.7429	.7670	.9994	.7253	.6712	.776
49	.6498	.6500	.6500	.6500	.6500	.6500	.6500	.6814	.6500	.6500	.6500	.6500	.6500	.6500	.6500	.655
50	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.000
51	.9196	.4761	.8407	1.0000	.8305	.5897	1.0000	1.0000	.9490	.9998	.9506	.6395	.8274	.9979	.9998	.881
52	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.967
53	.4000	.6611	.4005	.3909	.3919	.3214	.3996	.4004	.4005	.3909	.9993	.4000	.4004	.4005	.3887	.503
54	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.000
55	.8651	.8547	.8494	.9874	.7480	.8917	.8577	.9414	.9282	.9841	.9127	.9084	.9137	.9421	.9063	.890
56	.8000	.3000	.3000	.3093	.3051	.3052	.3000	.3000	.3000	.3000	.3052	.3000	.3050	.3000	.3000	.286
57	.6500	.7922	.6500	.2388	.1000	.7039	.1000	.7582	.1971	.6500	.7076	.7424	.3326	.2046	.4268	.523
58	1.0000	1.0000	.9868	.9671	1.0000	1.0000	1.0000	.9976	1.0000	.9999	1.0000	.9967	.9992	1.0000	.9911	.986
59	1.0000	.0519	.0143	.0500	.0450	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.457
60	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.000
62	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.1900	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.973
64	.5123	.1000	.3998	.1000	.0948	.4015	.5301	.5498	.1000	.1000	.1000	.1000	.1000	.1000	.5499	.207
65	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.000
66	.7885	.5034	.7791	.8440	.7291	.7443	.6762	.7982	.6942	.7425	.7236	.5866	.7499	.6108	.5988	.735
67	1.0000	.8419	1.0000	1.0000	1.0000	.9935	1.0000	1.0000	1.0000	1.0000	.9893	1.0000	.9974	1.0000	1.0000	.970
68	.9993	.8993	.6560	.8998	.6511	.9000	.5500	.8980	.5500	.5500	.6611	.5500	.5500	.9985	.7013	.738
69	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.000
70	.9995	1.0000	1.0000	.9970	1.0000	.9999	1.0000	1.0000	.9975	.9928	.9121	.9838	.6500	1.0000	.71284	.949
71	1.0000	.1000	1.0000	.1000	.3262	1.0000	.1827	.1833	.1000	1.0000	.1000	1.0000	.1379	.1000	1.0000	.513
72	1.0000	.9026	.9948	.9150	.9942	.9919	.9937	.9994	.9876	.7080	.9999	.9763	.9958	.9969	.8713	.958
73	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.000
74	.8543	.7007	.6941	.7431	.4015	.4015	.4015	.4015	.9095	.8486	.4015	.4015	.6664	.6585	.4012	.608
75	8733	.6474	.6182	.6344	.6707	.6490	.8091	.6772	.8318	.6339	.7693	.6497	.7537	.6812	.7571	.679
76	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.000

TABLE 2. DETERMINISTIC MODEL				
Shot Number	Empirical Value	Average Model Value	Difference	Signed Ranks of Difference
43	.734	.719	.015	11
44	.145	.700	-.555	-29
45	1.000	1.000	0.000	0
46	1.000	1.000	0.000	0
47	.100	.116	-.016	-12
48	.900	.776	.124	22
49	.930	.655	.275	25
50	1.000	1.000	0.000	0
51	.145	.881	-.736	-31
52	1.000	.967	.033	16
53	.668	.503	.165	23
54	1.000	1.000	0.000	0
55	1.000	.890	.110	21
56	.905	.286	.619	30
57	.550	.523	.027	14.5
58	1.000	.986	.014	10
59	1.000	.457	.543	28
60	.050	1.000	-.950	-32
62	1.000	.973	.027	14.5
64	.100	.207	-.107	-20
65	1.000	1.000	0.000	0
66	.668	.735	-.067	-19
67	.953	.970	-.017	-13
68	1.000	.738	.262	24
69	1.000	1.000	0.000	0
70	1.000	.949	.051	18
71	1.000	.513	.487	27
72	1.000	.958	.042	17
73	1.000	1.000	0.000	0
74	.905	.608	.297	26
75	.668	.679	-.011	-9
76	1.000	1.000	0.000	0

$\sum$ Positive Ranks $= 327$

$\sum |$ Negative Ranks $| = 165$

Critical T-Values ($\alpha = 0.05$) $= 142$ (lower), 350 (upper)

Critical T-Values ($\alpha = 0.10$) $= 158$ (lower), 334 (upper) standard normal distribution as the number of non-zero differences tends to infinity. For our example we have 32 observations, eight zero differences, and one tie with two tied differences; therefore, E [T] = 246 and Var [T] = 2809. We can calculate critical values by evaluating the equation $T = 53\hat{z} + 246$, where $\hat{z}$ is the $\alpha/2$ percentile of the standard normal distribution. As shown at the bottom of Table 2, even at an α-level of 0.10 there is no basis for rejecting the null hypothesis.

Table 3 contains the results for the stochastic model. Recall that R_i is the rank of z_i in the i^{th} observation $(y_i^1, y_i^2, ..., y_i^M, z_i)$, and T is defined as the sum of the R_i's. Under the null hypothesis of a valid model, z_i has the same distribution as $y_i^1, y_i^2, ..., y_i^M$; and therefore, R_i is uniformly distributed over the values $1, 2, ..., M + 1$. Modifying the results of Lehmann[15] by incorporating the number of observations, we can calculate the expectation of the test statistic,

$$E\,[T] = \frac{1}{2}\,[N\,(M + 2)]\,, \tag{3}$$

and the variance of the test statistic,

$$\mathrm{Var}\,[T] = \frac{1}{12}\,[N\,M\,(M + 2)] - \frac{1}{12\,[M + 1]}\,[\sum_{i=1}^{N} \sum_{j=1}^{n_i} (d_{ij}^3 - d_{ij})]\,, \tag{4}$$

where N is the number of observations, M is the number of replications of the model, and d_{ij} represents the number of tied values for the j^{th} tie in the i^{th} observation with n_i different ties in the i^{th} observation. Then $T^* = (T - E\,[T])/\sqrt{\mathrm{Var}\,[T]}$ will have approximately a standard normal distribution. For our example we have 32 observations, 30 replications, and 51 instances of tied values with varying numbers of ties; in this case E [T] = 512 and Var [T] = 1521. We can again calculate critical values, this time by evaluating the equation $T = 39\hat{z} + 512$, where $\hat{z}$ is the $\alpha/2$ percentile of the standard normal distribution. As shown at the bottom of Table 3, there is insufficient evidence to reject the null hypothesis at an α-level of 0.05; however, at an α-level of 0.10, the null hypothesis would be rejected.

Since in neither case could the null hypothesis be rejected at an α-level of 0.05, we must be concerned with the possibility of a Type II error; that is, accepting an invalid model. Figures 2-9 demonstrate the power of these tests against an alternative consisting of a shift in the mean. Consider the deterministic case. Referring to Figure 3, we see that if F (the distribution of the differences between the model output and the empirical data) is uniform, then the power of this test is very good since the probability of rejection rises quickly as the parameter increases in value. Conversely, Figure 4 demonstrates that if F is Cauchy, then the power of the test is rather poor. Results for the stochastic case are analogous. Figure 7 shows that the power of this test is very good if F (the distribution of the model output) and G (the distribution of the empirical data) are both uniform. However, as seen in Figure 8, if F and G are both Cauchy, then the power of the test is again rather poor.

Reynolds and Deaton[13] have proposed other test statistics more powerful against different alternatives; but for the loss of function data where empirical results that are close to the value one tend to be assigned that value, a shift in the mean seems to be an

TABLE 3. STOCHASTIC MODEL		
Shot Number	Empirical Value	Rank within Model Values
43	.734	16
44	.145	11
45	1.000	16
46	1.000	16
47	.100	8
48	.900	27
49	.930	31
50	1.000	16
51	.145	1
52	1.000	16
53	.668	27
54	1.000	16
55	1.000	31
56	.905	31
57	.550	11
58	1.000	22.5
59	1.000	24.5
60	.050	1
62	1.000	16.5
64	.100	13.5
65	1.000	16
66	.668	6
67	.953	[illegible].5
68	1.000	31
69	1.000	16
70	1.000	24
71	1.000	24.5
72	1.000	30
73	1.000	16
74	.905	30
75	.668	15
76	1.000	16

$\sum$ Ranks = 584

Critical T-Values ($\alpha = 0.05$) = 435 (lower), 589 (upper)

Critical T-Values ($\alpha = 0.10$) = 447 (lower), 577 (upper) appropriate alternative hypothesis. Since the power against this particular alternative is fairly good overall, our confidence in the hypothesis tests tends to increase. However, we would like to be able to make these tests and other tests still more powerful and, in the future, will be exploring methods to accomplish this.

V. SUMMARY

When referring to computer simulation models, a few authors continue to use the words verification and validation interchangeably; however, most distinguish between the two terms. Verification of a computer model assures that the simulation is behaving as the modeler intends, while validation assures that the simulation is behaving as the real world does. Verification is the process of debugging a computer program; validation is making it consistent with reality.

Prior to 1967 very little was written concerning the validation of simulations; but much has appeared since then, and there has been general agreement on several points - the most important being that to validate a computer simulation model, empirical observations are necessary and statistical tests are desirable. All validation techniques can be placed into one of five categories: judgemental comparisons, hypothesis testing, spectral analysis, sensitivity analysis, and indices of performance.

Nonparametric ranking techniques are one class of statistical hypothesis tests. We have advocated the Wilcoxon signed-ranks test as a validation procedure for deterministic simulation models and a combination of independent Mann-Whitney tests as a validation procedure for stochastic simulation models. They are statistical tests which assess empirical data to provide a certain level of confidence in the computer model. The main disadvantage of both is the same as that of all hypothesis testing techniques; namely, their concern for protecting against Type I errors, sometimes at the expense of Type II errors. A Type I error results in rejecting a valid simulation model - unfortunate, but not as potentially dangerous as accepting an invalid simulation model, which is known as a Type II error. For any particular test we can get an indication of the probability of a Type II error by generating a series of curves that will allow us to examine the power of the test against various alternatives.

Power is defined as the probability of rejecting a false null hypothesis, and we would like this value to be as close to one as possible. For our advocated tests we have evaluated the power for some specific alternative hypotheses by incorporating a Monte-Carlo procedure into a computer program, which allowed us to perform thousands of replications. Each replication represents a case in which the alternative hypothesis was true, and we determined whether or not the test rejected the null hypothesis. Obviously, we can not compute power against an alternative hypothesis as general as, "The simulation model is invalid." However, in being more specific we are forced to examine an array of different alternative hypotheses; and while a test may be powerful against a subset of these alternatives (such as a shift in the mean of a distribution), it might be less so against others. The most we can hope for is reasonable power against alternatives important to a particular investigation. Both the Wilcoxon signed-ranks test and the combination of independent Mann-Whitney tests appear to have reasonable power against a shift in the mean, but we would like to be able to increase it.

For any given alternative hypothesis there are several ways of increasing the power. One such way can be seen in Figures 2-9 - increasing the number of observations. Another way is to reduce the level of confidence in the test itself; that is, allow the probability of a Type I error to increase. In the future we will be investigating other methods for increasing the power of statistical hypothesis tests in general and of the two we have advocated in particular. These methods will include a statistical procedure known as bootstrapping, a mathematical theory known as fuzzy sets, and, possibly, a combination of the two. Because of the importance in this area of computer simulation validation, we hope to develop ways to make these tests more powerful against a wide range of alternatives while still permitting them to provide acceptable levels of confidence in their results.

ACKNOWLEDGMENT

The authors acknowledge and appreciate the assistance of Larry Losie, who took time to explain details of the VAST simulation model and who made other substantial efforts which culminated in most of the data appearing in Tables 1-3 of this report.

REFERENCES

1. Fishman, G.S. and Kiviat, P.J., "Digital Computer Simulation: Statistical Considerations," Memorandum RM-5387-PR, The Rand Corporation, 1967.

2. Winter, E.M., Wisemiller, D.P., and Ujihara, J.K., "Verification and Validation of Engineering Simulations with Minimal Data," Proceedings of the 1976 Summer Computer Simulation Conference, 1976.

3. Stone, M., "Cross-Validating Choice and Assessment of Statistical Prediction," Journal of the Royal Statistical Society, Series B-36, 1974

4. Van Horn, R., "Validation," The Design of Computer Simulation Experiments, Duke University Press, 1969.

5. Naylor, T.H. and Finger, J.M., "Verification of Computer Simulation Models," Management Science, Vol. 14 No. 2, 1967.

6. Law, A. M., Simulation Modeling and Analysis, University of Wisconsin, 1979.

7. Wright, R.D., "Validating Dynamic Models: An Evaluation of Tests of Predictive Power," Proceedings of the 1972 Summer Computer Simulation Conference, 1972.

8. Baird, A.M., Goldman, R.B., Bryan, W.C., Holt, W.C., and Belrose, F.M., "Verification and Validation of RF-Environmental Models - Methodology Overview," Boeing Aerospace Company, 1980.

9. Tytula, T.P., "A Method for Validating Missile System Simulation Models," Technical Report E-78-11, U.S. Army Missile Research and Development Command, 1978.

10. Garrett, M., "Statistical Validation of Simulation Models," Proceedings of the 1974 Summer Computer Simulation Conference, 1974.

11. Gilmour, P., "A General Validation Procedure for Computer Simulation Models," The Austrailian Computer Journal, Vol. 5 No. 3, 1973.

12. Sargent, R.G., "Developing Statistical and Cost-Risk Procedures for Validation of Simulation Models," US. Army Research Office Proposal Number 18201-M, 1980.

13. Reynolds, M.R. and Deaton, M.L., "Comparisons of Some Tests for Validation of Stochastic Simulation Models," Commun. Statist. - Simula. Computa., Vol. 11 No. 6, 1982.

14. Conover, W.J., Practical Nonparametric Statistics, John Wiley & Sons, Inc. 1971.

15. Lehmann, E.L., Nonparametrics: Statistical Methods Based on Ranks, Holden-Day, Inc., 1975.

16. Mihram, G.A., Simulation: Statistical Foundations and Methodology, Academic Press, Inc., 1972.

17. Van Elteren, P., "On the Combination of Independent Two Sample Tests of Wilcoxon," Bulletin de l'institute International de Statistique, 37, 1960.

18. Reynolds, M.R., Burkhart, H.E., and Daniels, R.F., "Procedures for Statistical Validation of Stochastic Simulation Models," Forest Science, Vol. 27 No. 2, 1981.

19. Hafer, T.F. and Hafer, A.S., "Vulnerability Analysis for Surface Targets (VAST): An Internal Point-Burst Vulnerability Model," ARBRL-TR-02154, U.S. Army Ballistic Research Laboratory, 1979.

DISTRIBUTION LIST

No. of Copies — Organization

12 Administrator
Defense Tech Info Ctr
ATTN: DTIC-DDA
Cameron Station
Alexandria,
VA 22304-6145

1 Director
Inst for Def Analyses
1818 Beauregard St.
Alexandria, VA 22311

1 Director
Defense Advanced Research
Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209

1 HQDA (DAMA-ART-M)
Washington, DC 20310

1 HQDA (DAMA-ARR)
Washington,
DC 20310-0632

1 HQDA (DACA-CW)
Washington, DC 20310

1 HQDA (DAMI)
Washington, DC 20310

1 Director
US Army Engineer Waterways Experiment Station
P. O. Box 631
Vicksburg, MS 39108

1 Commander
US Army Materiel Command
ATTN: AMCDRA-ST
5001 Eisenhower Avenue
Alexandria, VA 22333

No. of Copies — Organization

1 Commander
US Army Materiel Command
ATTN: AMCLD
5001 Eisenhower Avenue
Alexandria,
VA 22333-0001

1 Commander
Armament R&D Center
US Army AMCCOM
ATTN: SMCAR-TDC
Dover, NJ 07801

1 Commander
Armament R&D Center
US Army AMCCOM
ATTN: SMCAR-TSS
Dover, NJ 07801

2 Commander
US Army Armament Research
and Development Command
ATTN: SMCAR-SC-Y,
Mr. Gaydos
SMCAR-LC-A,
Mr. Brooks
Dover, NJ 07801

4 Commander
Armamment R&D Center
US Army AMCCOM
ATTN: SMCAR-LC
SMCAR-SE
SMCAR-SA
SMCAR-AC
Dover, NJ 07801

1 Commander
US Army Armament,Munition
and Chemical Command
ATTN: SMCAR-ESP-L
Rock Island, IL 61299

10 CENTRAL INTELLIGENCE AGENCY
OFFICE OF CENTRAL REFERENCE
DISSEMINATION BRANCH
ROOM GE-47 HQS
WASHINGTON, D.C. 20502

DISTRIBUTION LIST

No. of
Copies Organization

1 Commander
US Army Armament, Munitions
and Chemical Command
ATTN: AMSAR-SA,
Mr. Michels
Rock Island, IL 61299

1 Director
Benet Weapons Laboratory
Armament R&D Center
US Army AMCCOM
ATTN: SMCAR-LCB-TL
Watervliet, NY 12189

1 Commander
US Army Aviation Research
and Development Command
ATTN: AMSAV-E
4300 Goodfellow Blvd
St. Louis, MO 63120

1 Commander
US Army Air Mobility
R&D Laboratory
Ames Research Center
Moffett Field, CA 94035

1 Director
Appl. Tech. Directorate
USAARTA
(AVSCOM)
ATTN: DAVDL-EU-SY-RPV
Fort Eustis, VA 23604

1 Commander
US Army Troop Support and
Aviation Materiel
Readiness Command
ATTN: AMSTS-G
4300 Goodfellow Boulevard
St. Louis, MO 63120

No. of
Copies Organization

1 Commander
US Army Communications-
Electronics Command
ATTN: AMSEL-ED
Fort Monmouth, NJ 07703

1 Commander
US Army Communications
Command
ATTN: ATSI-CD-MD
Fort Huachuca, AZ 85613

1 Commander
ERADCOM Tech. Library
ATTN: DELSD-L(Rpts. Sec)
Fort Monmouth,
NJ 07703-5301

1 Commander
US Army Missile Command
ATTN: AMSMI-R
Redstone Arsenal,
AL 35898

1 Commander
US Army Missile Command
ATTN: AMSMI-YDL
Redstone Arsenal,
AL 35898

1 Commander
US Army Belvoir R&D
Center
ATTN: AMDME-WC
Fort Belvoir,
VA 22060-5606

1 Commander
US Army Tank Automotive
Command
ATTN: AMSTA-TSL
Warren, MI 48090

DISTRIBUTION LIST

No. of Copies Organization

2 Commander
US Army Research Office
ATTN: AMXRO-MA,
Dr. J. Chandra
Dr. R. Launer
P. O. Box 12211
Research Triangle Park,
NC 27709-2211

1 Commander
US Army Combat Dev &
Experimentation Command
ATTN: ATEC-SA,
Dr. M. R. Bryson
Fort Ord, CA 93941

1 Commander
US Army Harry Diamond
Laboratory
ATTN: DELHD-RT-RD,
Mr. R. Antony
2800 Powder Mill Rd
Adelphi, MD 20783

1 Commander
US Army Logistics Center
ATTN: ATLC-OOM,
J. Knaub
Ft. Lee, VA 23801

1 Director
US Army Concepts Analysis
Agency
ATTN: CSCA-AST,
Mr. C. Bates
8120 Woodmont Avenue
Bethesda, MD 20014

1 Director
USA Res, Dev and
Standardization Group (UK
ATTN: Dr. J. Gault
Box 65
APO New York, NY 09510

No. of Copies Organization

1 Director
US Army Concepts Analysis
Agency
ATTN: CSCA-AST,
Mr. B. Graham
8120 Woodmont Avenue
Bethesda, MD 20014

1 Director
Walter Reed Army
Institute of Research
Walter Reed Army
Medical Center
ATTN: SGRD-UWE,
Dr. D. Tang
Washington, DC 20012

1 President
US Army Airborne,
Electronics & Special
Warfare Board
Fort Bragg, NC 28307

1 President
US Army Armor &
Engineer Board
Fort Knox, KY 40121

1 President
US Army Artillery Board
Fort Sill, OK 73503

1 AFWL/SUL
Kirtland AFB,
NM 87117-6008

1 AFWL/NTES(Dr. Ross)
Kirtland AFB,
NM 87117-6008

1 Commander
US Army Materials and
Mechanics Research Ctr
Watertown, MA 02172

DISTRIBUTION LIST

No. of Copies Organization

1 Commander
US Army Training and
Doctrine Command
Fort Monroe, VA 23651

1 Commander
US Army TRADOC Systems
Analysis Activity
ATTN: ATAA-SL, Tech Lib
White Sands Missile Range
NM 88002

2 Commandant
US Army Armor School
ATTN: Armor Agency
ATSB-CD-MM
Fort Knox, KY 40121

1 Commandant
US Army Artillery School
ATTN: ATSF-CA,
Mr. Minton
Fort Sill, OK, 73503

1 Commandant
US Army Aviation School
ATTN: Aviation Agency
Fort Rucker, AL 36360

1 Commandant
US Army Infantry School
ATTN: ATSH-CD-CSO-OR
Fort Benning, GA 31905

1 Commandant
US Army Infantry School
ATTN: ATSH-CD-CSO-OR,
Mr. J. D'Errico
Fort Benning, GA 31905

1 Commandant
US Army Intelligence Sch
ATTN: Intel Agcy
Fort Huachuca, AZ 85613

No. of Copies Organization

1 Commander
US Army Development &
Employment Agency
ATTN: MODE-TED-SAB
Fort Lewis, WA 98433

1 Chief of Naval Operations
ATTN: OP-721
Department of the Navy
Washington, DC 20350

1 Chief of Naval Materiel
ATTN: MAT-0324
Department of the Navy
Washington, DC 20360

2 Commander
Naval Air Systems Command
ATTN: WEPS,
Mr. R. Sawyer
AIR-604
Washington, DC 20360

1 Commander
Naval Air Development
Center, Johnsville
ATTN: Code SRS
Warminster, PA 18974

2 Commander
Naval Surface Weapons Ctr
ATTN: DX-21, Lib Br.
Mr. N. Ruppert
Dahlgren, VA 22448

2 Commander
Naval Surface Weapons Cen
ATTN: Code G11,
Mr. Ferrebee
Code G12,
Mr. Hornbaker
Dahlgren, VA 22448

DISTRIBUTION LIST

No. of Copies	Organization
3	Commander Naval Weapons Center ATTN: Code 31804 Code 3835 Code 338 China Lake, CA 93555
1	Commander Naval Research Lab Washington, DC 20375
1	Commander David Taylor Naval Ships Research & Development Center ATTN: Tech Library Bethesda, MD 20084
1	Commandant US Marine Corps ATTN: AAW-1B Washington, DC 20380
1	Commandant US Marine Corps ATTN: POM Washington, DC 20380
1	Commanding General Fleet Marine Force, Atlantic ATTN: G-4 (NSAP) Norfolk, Va 23511
1	Commander Marine Corps Development and Education Command (MCDEC) Quantico, VA 22134
1	HQ USAF/SAMI Washington, DC 20330-5425
3	AFSC (SCFO; SDW; DLCAW) Andrews AFB, MD 20331
2	ADTC (DLODL; ADBRL-2) Eglin AFB,.FL 32542
1	AFATL (DLMM) Eglin AFB, FL 32542
1	USAFTAWC/ADTC Eglin AFB, FL 32542
1	Air Force Armament Lab ATTN: AFATL/DLODL Eglin AFB, FL 32542-5000
1	TAC (INAT) Langley AFB, VA 23665
1	AFWAL/FIBC Wright-Patterson AFB, OH 45433
1	FTD (ETD) Wright-Patterson AFB, OH 45433
1	Battelle Columbus Laboratories ATTN: Ordnance Div 505 King Avenue Columbus, OH 43201
1	Zernow Tech Services 425 W. Bonita Ave Suite 208 San Dimas, CA 91773

DISTRIBUTION LIST

No. of Copies Organization

2 Southwest Research Inst
Dept of Mech Sciences
ATTN: Mr. A. Wenzel
Mr. P. Zabel
8500 Culebra Road
San Antonio, TX 78284

1 Oklahoma State University
Field Office
ATTN: Mrs. Ann Peebles
P. O. Box 1925
Eglin Air Force Base,
FL 32542

1 Prof. Lotfi Zadeh
Dept of Electrical Eng
& Comp Science
University of California
Berkeley, CA 94720

1 Prof. James T.P. Yao
School of Civil Eng
Civil Eng Bldg
Purdue University
West Lafayette, In 47907

1 Dr. William H. Friedman
Dept. Econ & Dec. Sci.
Loyola College
4501 N. Charles St
Baltimore, MD 21210-2699

1 Prof. J.L.Chameau
Geotech Eng
Grissom Hall
Purdue University
West Lafayette, IN 47907

1 Machine Intelligence Inst
Iona College
ATTN: Dr.R. Yager
New Rochelle, NY 10801

No. of Copies Organization

1 Dr. Felix S. Wong
Weidlinger Associates
620 Hansen Way
Suite 100
Palo Alto,.CA 94304

1 Dr. Steven B. Boswell
MIT Lincoln Lab, Group 94
Lexington, MA 02173-0073

Aberdeen Proving Ground

12 Dir, USAMSAA
ATTN: AMXSY-D,
Mr. K. Myers
AMXSY-MP,
Mr. H. Cohen
AMXSY-A,
Mr. D. O'Neill
AMXSY-RA,
Mr. R. Scungio
AMXSY-GS,
Mrs. M. Ritondo
Dr. M. Starks
AMXSY-AAG,
Mr. W. Nicholson
Mr. C. Abel
AMXSY-G
Mr. J. Kramar
AMXSY-J,
Mr. J. Blomquist
Mr. J. Matts
AMXSY-LR,
Mr. W. Webster

1 Cdr, USATECOM
ATTN: AMSTE-TO-F

3 Cdr, CRDC, AMCCOM
ATTN: SMCCR-RSP-A
SMCCR-MU
SMCCR-SPS-IL

USER EVALUATION SHEET/CHANGE OF ADDRESS

This Laboratory undertakes a continuing effort to improve the quality of the reports it publishes. Your comments/answers to the items/questions below will aid us in our efforts.

1. BRL Report Number ______________________ Date of Report __________

2. Date Report Received ______________________________

3. Does this report satisfy a need? (Comment on purpose, related project, or other area of interest for which the report will be used.) ______________

4. How specifically, is the report being used? (Information source, design data, procedure, source of ideas, etc.) ______________

5. Has the information in this report led to any quantitative savings as far as man-hours or dollars saved, operating costs avoided or efficiencies achieved, etc? If so, please elaborate. ______________

6. General Comments. What do you think should be changed to improve future reports? (Indicate changes to organization, technical content, format, etc.)

CURRENT ADDRESS

____
Name

____
Organization

____
Address

____
City, State, Zip

7. If indicating a Change of Address or Address Correction, please provide the New or Correct Address in Block 6 above and the Old or Incorrect address below.

OLD ADDRESS

____
Name

____
Organization

____
Address

____
City, State, Zip

(Remove this sheet along the perforation, fold as indicated, staple or tape closed, and mail.)